



REGIONAL  
CENTRE  
NETWORK

# SRCNet update for NREN forum

Rosie Bolton

Head of Data Operations

26/02/2023



# Staged Delivery Strategy

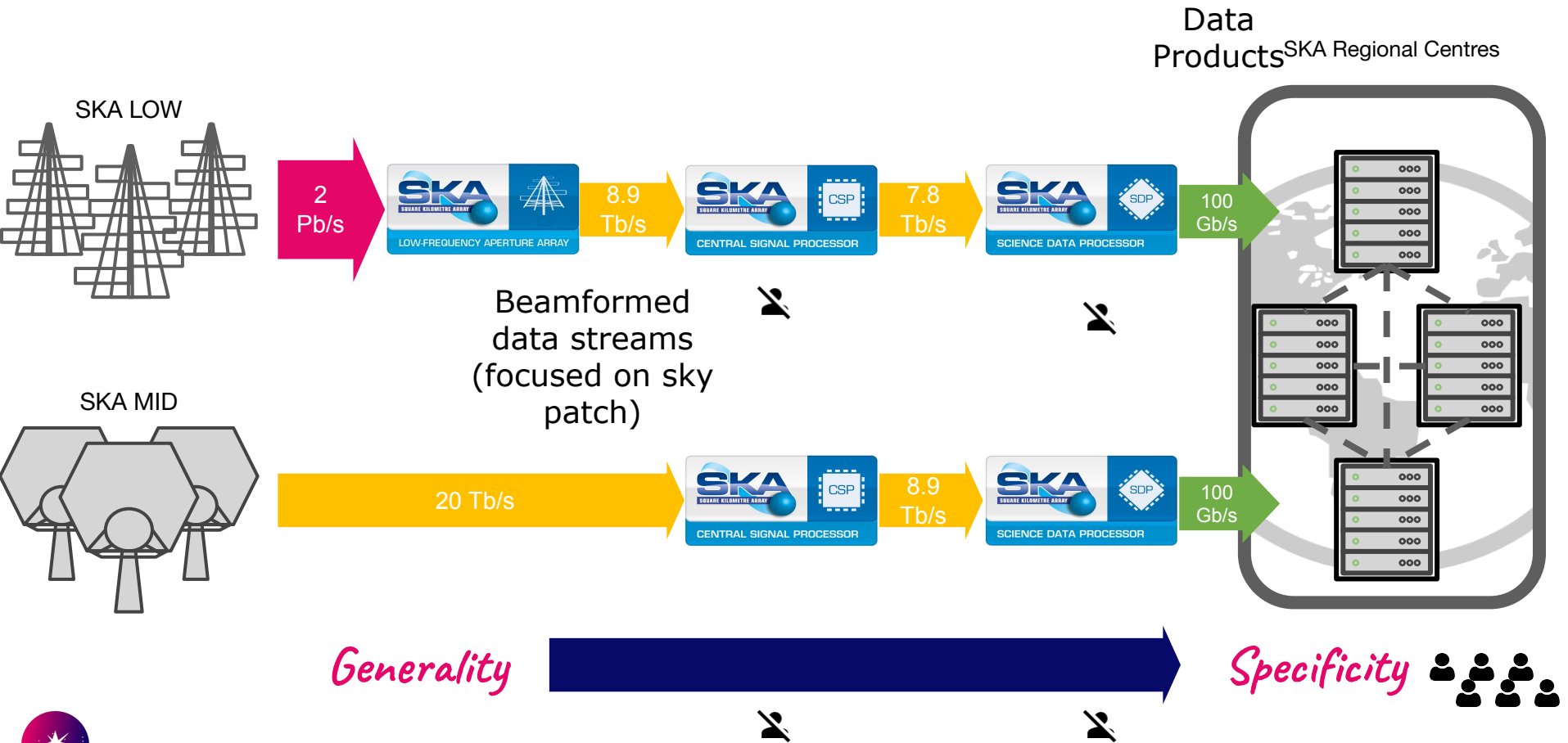
Milestone event (earliest)		SKA-Mid (end date)	SKA-Low (end date)
AA0.5	4 dishes 6 stations	2025 May	2024 Nov
AA1	8 dishes 18 stations	2026 May	2025 Nov
<b>AA2</b>	<b>64 dishes</b> <b>64 stations</b>	<b>2027 Apr</b>	<b>2026 Dec</b>
AA*	144 dishes 307 stations	2028 Jan	2028 Mar
Operations Readiness Review		2028 Apr	2028 Aug
AA4	197 dishes 512 stations	TBD	TBD

**See Joe's talk for best possible version of this timeline**

- **Target: Build the SKA Baseline design (197 Mid Dishes, 512 Low Stations = "AA4")**
- Not all funding yet secured
- Develop the earliest possible working demonstration of the architecture and supply chain (AA0.5)
- Maintain a continuously working and expanding facility that demonstrates the full performance capabilities of the SKA Design
- User interaction with data products Expected to start towards the end of AA2 when the SKA is scientifically interesting (**Science Verification**)
- SRCs expected to be involved in this



# SKA Data Flow (rates for AA4)



# Estimates of SKA Data Output... why this is hard!

The range of SKA science is vast - huge differences in the data volumes required

Examples here are taken from the parametric model used to predict computing needs *within* SKAO, but they also include output data rates

The specific examples are quite out of date - take home message is the range - from <1 to 100 Gbit/s data generation rate, and that the fractions of time assumed greatly impact overall rates

SKA1\_Low:

HPSO	Time [%]	Tobs [h]	Npix (side)	Channels (DPrepB)	Channels (DPrepC)	Image size [GB]	Non-Vis Rate [Gbit/s]	Visibility Size [TB]	Visibility Rate [Gbit/s]	Total Rate [Gbit/s]
hps001	15.6	5.00	18344	500	1500	2.7	8.5	205.8	91.4	99.9
hps002a	15.6	5.00	18344	500	1500	2.7	8.5	205.8	91.4	99.9
hps002b	15.6	5.00	18344	500	1500	2.7	8.5	205.8	91.4	99.9
hps004a	39.8	0.67	-	-	-	-	0.7	-	-	0.7
hps005a	13.4	0.67	-	-	-	-	2.6	-	-	2.6
<b>Average</b>	-	-	-	-	-	-	4	-	42.8	47.4

SKA1\_Mid:

HPSO	Time [%]	Tobs [h]	Npix (side)	Channels (DPrepB)	Channels (DPrepC)	Image size [GB]	Non-Vis Rate [Gbit/s]	Visibility Size [TB]	Visibility Rate [Gbit/s]	Total Rate [Gbit/s]
hps004b	1.0	0.17	-	-	-	-	2.3	-	-	2.3
hps004c	3.1	0.17	-	-	-	-	2.3	-	-	2.3
hps005b	2.1	0.25	-	-	-	-	6.9	-	-	6.9
hps013	6.5	8.00	18344	160	3200	5.1	4.2	-	-	4.2
hps014	2.6	8.00	18814	300	5000	2.8	2.8	-	-	2.8
hps015	16.5	4.40	10837	260	2500	0.9	0.8	-	-	0.8
hps018	13.1	0.02	-	-	-	-	0.1	-	-	0.1
hps022	7.9	8.00	110601	1000	0	97.9	48.1	-	-	48.1
hps027and33	13.1	0.12	23549	700	0	4.4	99.3	-	-	99.3
hps032	13.1	2.20	-	-	-	-	1.3	-	-	1.3
hps037a	13.1	3.80	94195	700	0	71.0	60.6	-	-	60.6
hps037b	2.6	8.00	94195	700	0	71.0	28.8	-	-	28.8
hps037c	2.6	8.00	94195	700	0	71.0	28.8	-	-	28.8
hps038a	1.3	8.00	113204	1000	0	102.5	50.4	-	-	50.4
hps038b	1.3	8.00	113204	1000	0	102.5	50.4	-	-	50.4
<b>Average</b>	-	-	-	-	-	-	28.4	-	0.0	28.4

Old slide - you've seen this before!



SKAO Science  
Working  
Groups

# Year in the Life: AA\* / AA4

Cosmology

Cradle of Life

Epoch of  
Reionization

Extragalactic  
Continuum

Extragalactic  
Spectral Line

Gravitational  
Waves

High Energy  
Cosmic Particles

HI Galaxy  
Science

Magnetism

Our Galaxy

Pulsars

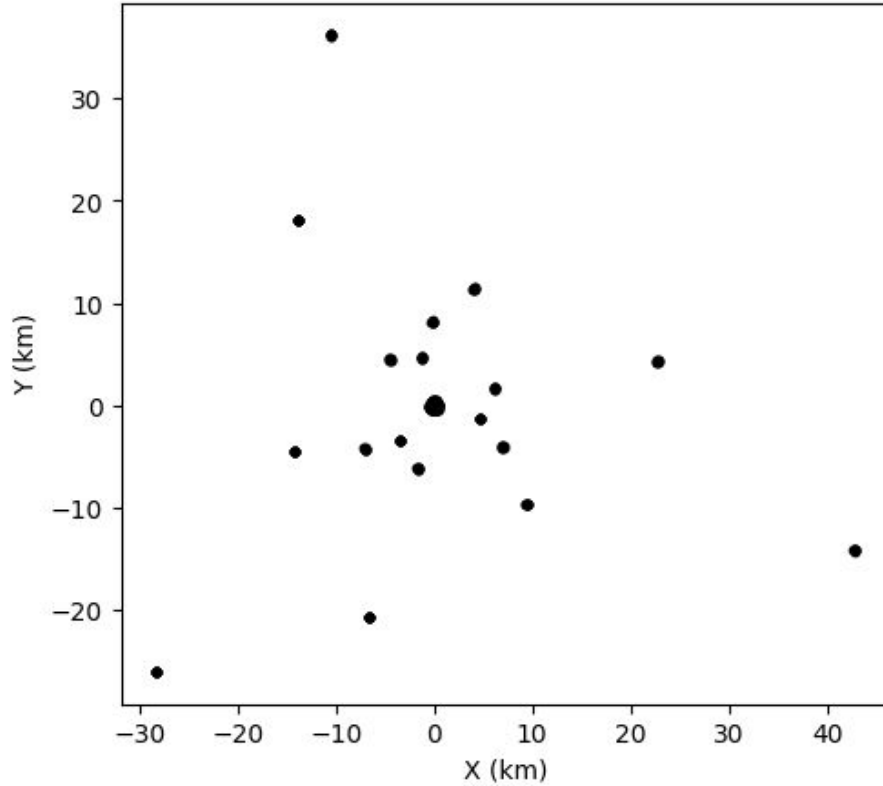
Solar,  
Heliospheric &  
Ionospheric  
Physics

Transients

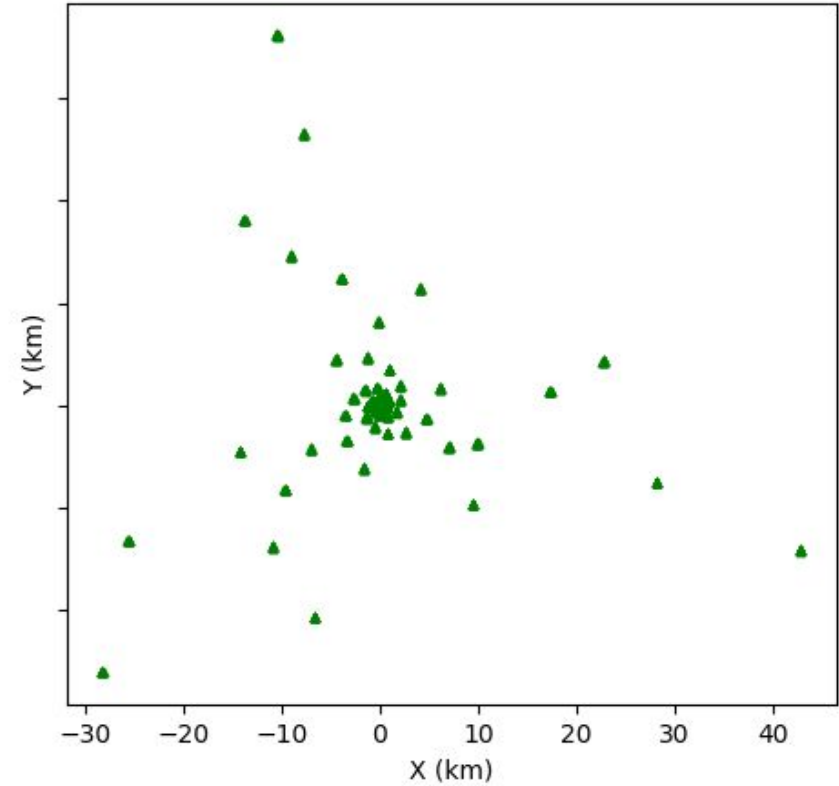
VLBI

# Staged delivery examples: LOW (example only)

LOW AA\*

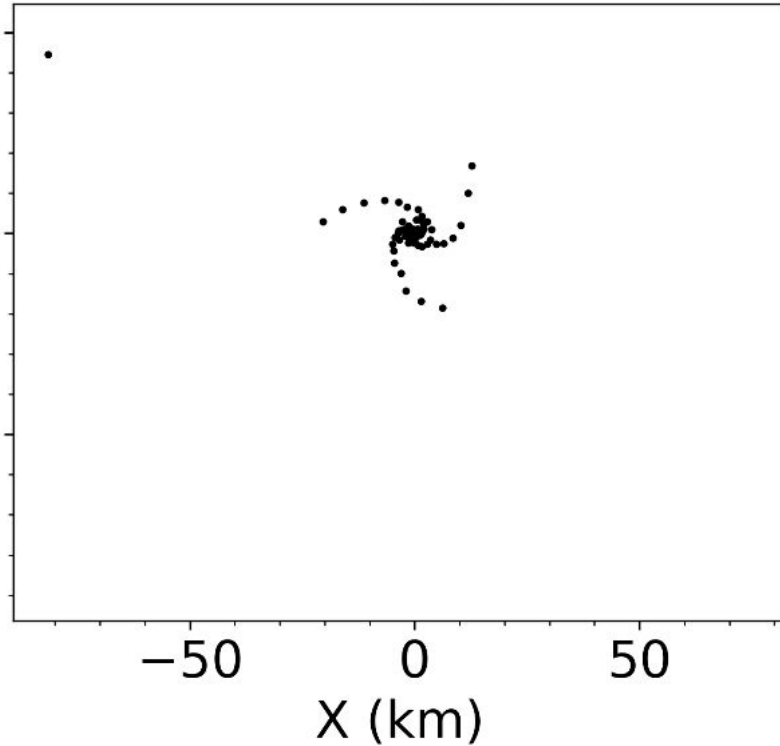


LOW AA4

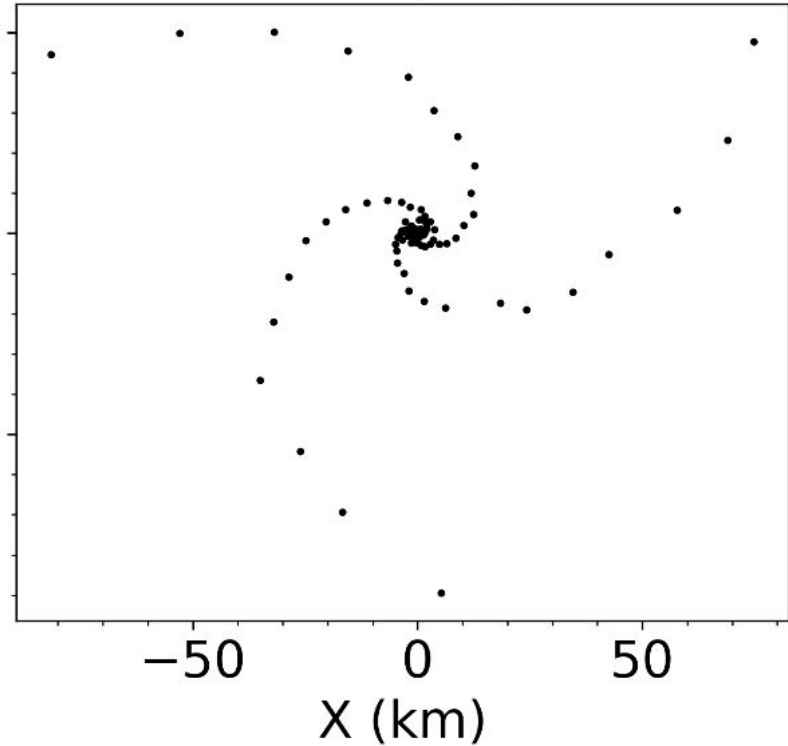


# Staged delivery examples: MID (example only)

AA\* array layout



AA4 array layout



# Year in the Life (led by Shari Breen, Head of Science Operations, SKAO)

- Critical review of realistic science schedule for AA\* into AA4
- Scientifically feasible experiments
- Achievable observing patterns (many previous HPSOs assumed several thousand hours of observing - total schedule was over a decade)
- Consider roll-out of observing modes
- Will consider data products and give output data rates
- **Results by April GEANT NGN SIG meeting**



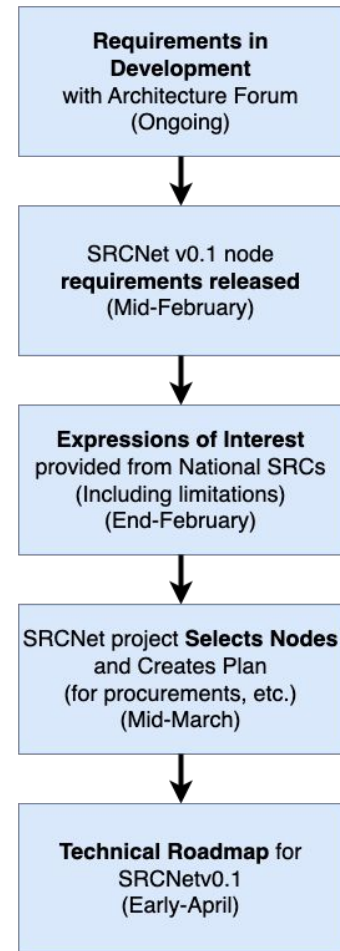


**But... also need to consider our technical scale-up roadmap**

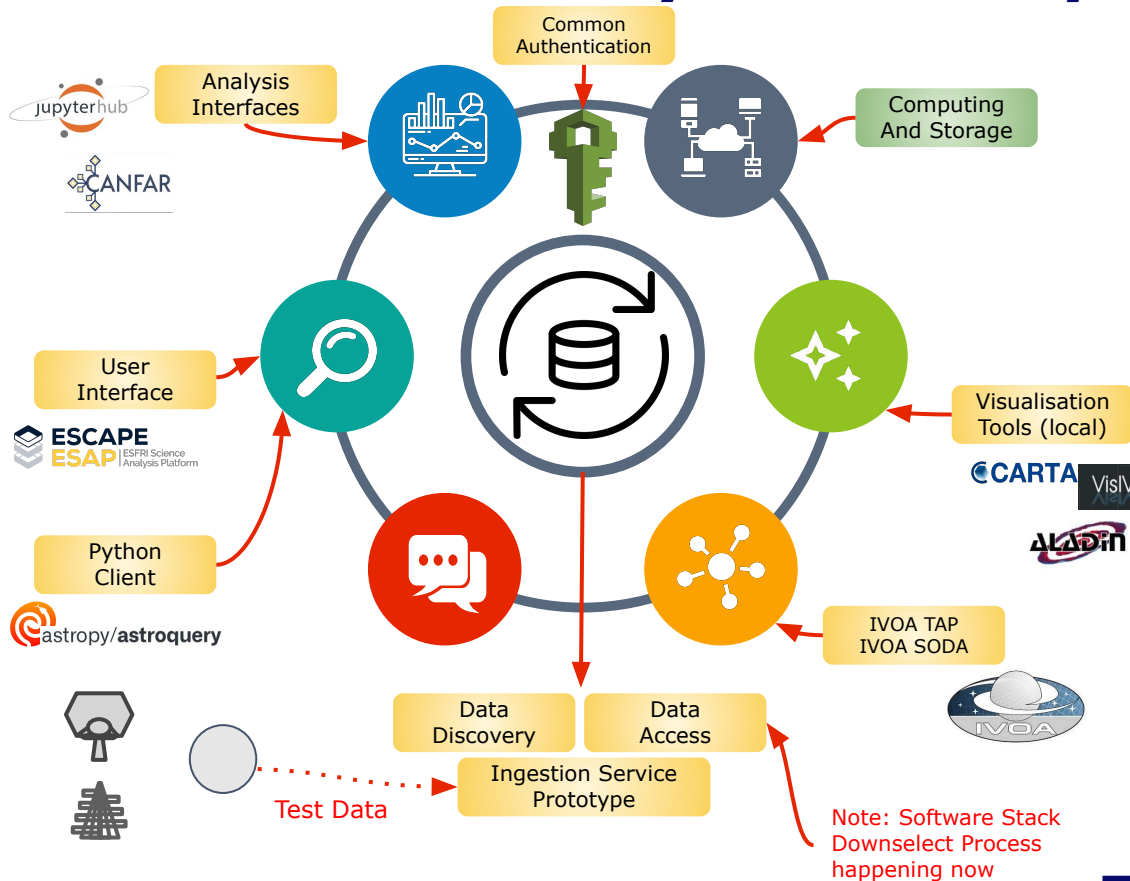


# SRCNet v0.1 deployment

- Deployment of first operational version of v0.1 foreseen for Nov-2024
- Stable and maintained limited numbers of nodes (~4) with enough human and hardware resources
- Document describing requirements to have a node is being developed
  - SRCs will reply with a possible expression of interest to be one node, including available resources
  - Deployment plan to be produced for early-March



# Basic Functionality Covered by v0.1



- Common Authentication
  - IAM
- Visualisation Tools (local)
- IVOA Protocols
  - TAP, SODA
- Data Discovery and Access from Data Lake
- Ingestion Service Prototype
- Python Client
  - Astroquery Module
- User Interface
  - ESAP
  - <https://esap.srcdev.skao.int/>
- Analysis Interfaces
  - JupyterHub
  - CANFAR Science Platform

Target ~20PB storage

# SKA Regional Centre Capabilities

Archive of data products continues to grow over the 50yr+ lifetime of the Observatory.

Eventually similar in scale to LHC / HL-LHC



**Data Management**

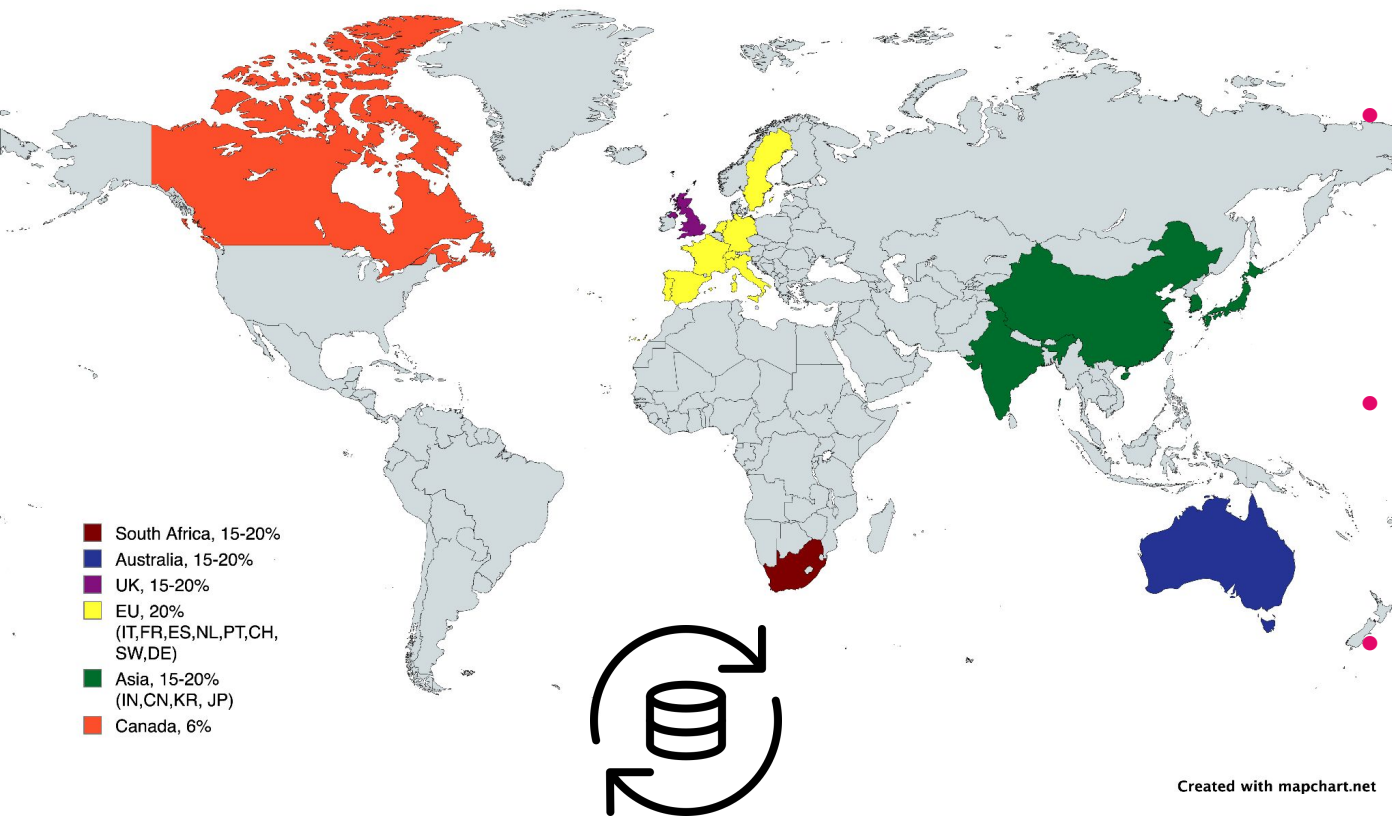
Dissemination of Data to SRCs and Distributed Data Storage

Distributed data management is at the heart of SRCNet

appropriately managing replicas to avoid duplication is the best way to achieve long term cost savings



# SKA Regional Centre Broad Distribution: Fair Share



- Roughly, 6 global zones of equivalent size (Canada smaller)
- Distribute two base copies of each data product to different countries, and perhaps insist to different regions
- Average incoming rate per region not more than  $2 \times 20 \text{ Gbit/s} = 40 \text{ Gbit/s}$  ( $\sim 2 \times 6 \text{ Gbit/s}$  for Canada)
- Max 100 Gbit/s out of SA and AUS

Created with mapchart.net



# Summary notes

- Assumption of 100 Gbit/s is old
  - based on previous mix of data products with old assumptions for telescope and science data parameters
  - Is actually dominated by "raw" data
- An updated year in the life model is the best way to estimate changes in data rate
  - **cannot assume that data rate will go down significantly** (may need 10x higher fraction of raw data, for example)
  - This taking some time, because we are working hard to get it "right" - misinformation more damaging than no information!
- Technical testing requirements will dominate in early years - SRCNet0.1 means target 20PBytes storage in 2025
- **Initial SRCNet sites will be confirmed in March 2024, up and running by end of year.**
- **Distributed Data Management service in SRCNet will orchestrate data distribution globally**



# SKA Network @ LHCOPN/LHCONE

<https://indico.cern.ch/event/1349135/>

## Half day session, topics could include

- SKAO construction update
- SRCNet development update: decisions taken, what are the SRCNet0.1 sites, and connectivity required / actual connectivity
- Data flow models / data management lifecycle
- Data moving plans for SRCNet0.1
- Technical working group update on Network requirements
- Approaches for network capacity development (e.g. GREN, how WLCG did things and why)

For discussion today!



